



Experiences Developing a Semantic Representation of Product Quality, Bias, and Uncertainty for a Satellite Data Product

Patrick West¹, Gregory Leptoukh², Stephan Zednik¹, Chris Lynnes², Suraiya Ahmad³, Jianfu Pan⁴, Peter Fox¹

- (1) Tetherless World Constellation
- (2) NASA Goddard Space Flight Center
- (3) NASA Goddard Space Flight Center/Innovim
- (4) NASA Goddard Space Flight Center/Adnet Systems, Inc.





Issue

- Climate model and various environmental monitoring and protection applications have begun to increasingly rely on satellite measurements.
- Research application users seek good quality satellite data, with uncertainties and biases provided for each data point
- Remote-sensing quality issues are addressed rather inconsistently and differently by different communities.



Where are we in respect to this data challenge?

*“The user cannot **find** the data;
If he can find it, cannot **access** it;
If he can access it, ;*

*he doesn't know **how good** they are;
if he finds them good, he can not **merge**
them with other data”*

The Users View of IT, NAS 1989



Challenges in dealing with Data Quality

- Q: Why now? What has changed?
- A: With the recent revolutionary progress in data systems, dealing with data from many different sensors finally has become a reality.

Only now, a systematic approach to remote sensing quality is on the table.

- NASA is beefing up efforts on data quality.
- ESA is seriously addressing these issues.
- QA4EO: an international effort to bring communities together on data quality.
- GeoVigue



Intercomparison of data from multiple sensors

Data from multiple sources to be used together:

- Current sensors/missions: MODIS, MISR, GOES, OMI.
- Future missions: ACE, NPP, JPSS, Geo-CAPE
- European and other countries' satellites
- Models

Harmonization needs:

- It is not sufficient just to have the data from different sensors and their provenances in one place
- Before comparing and fusing data, things need to be harmonized:
 - Metadata: terminology, standard fields, units, scale
 - Data: format, grid, spatial and temporal resolution, wavelength, etc.
 - Provenance: source, assumptions, algorithm, processing steps
 - **Quality: bias, uncertainty, fitness-for-purpose, validation**

Dangers of easy data access without proper assessment of the joint data usage - *It is easy to use data incorrectly*



Three projects with data quality flavor

- We have three projects where different aspects of data quality are addressed.
- We mostly deal with aerosol data
- I'll briefly describe them and then show why they are related



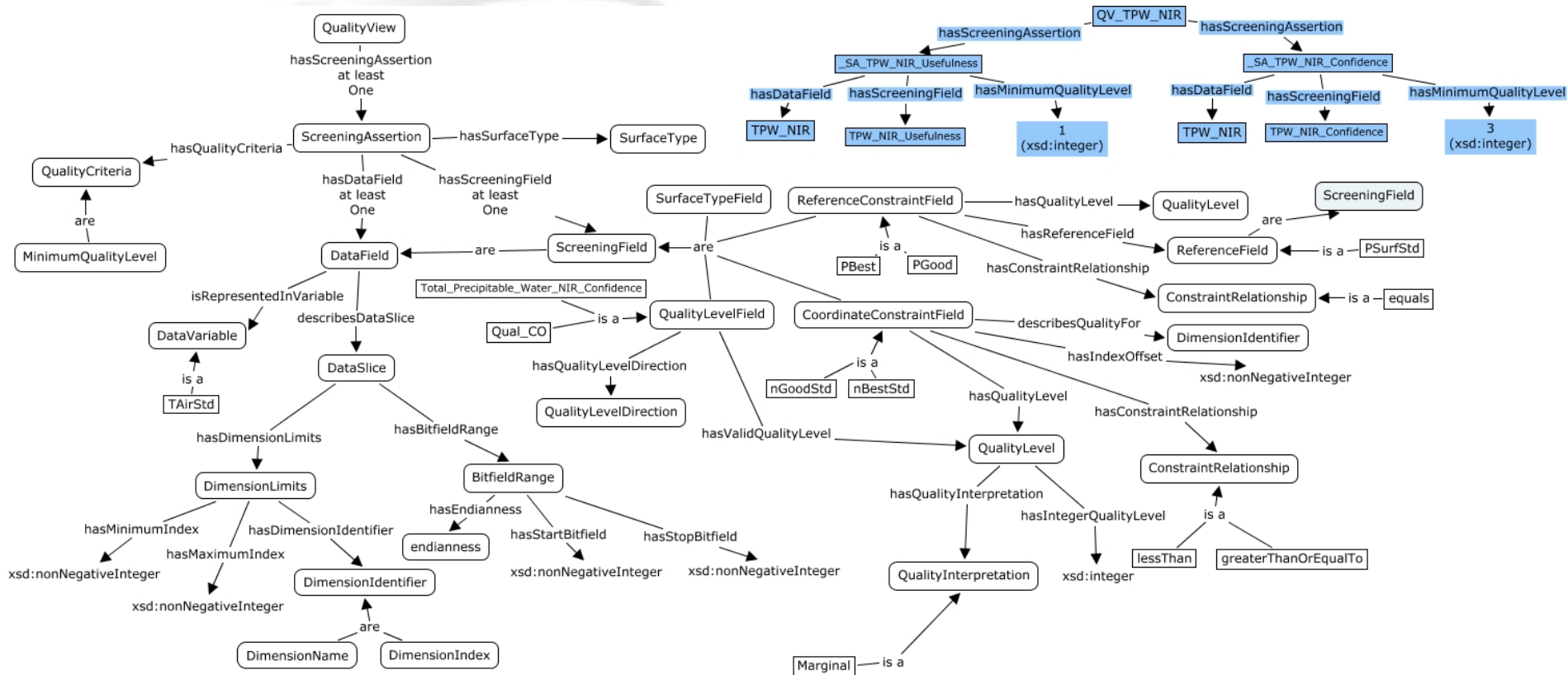
Data Quality Screening Service for Remote Sensing Data

The DQSS filters out bad pixels for the user

- Default user scenario
 - Search for data
 - Select science team recommendation for quality screening (filtering)
 - Download screened data
- More advanced scenario
 - Search for data
 - Select custom quality screening parameters
 - Download screened data

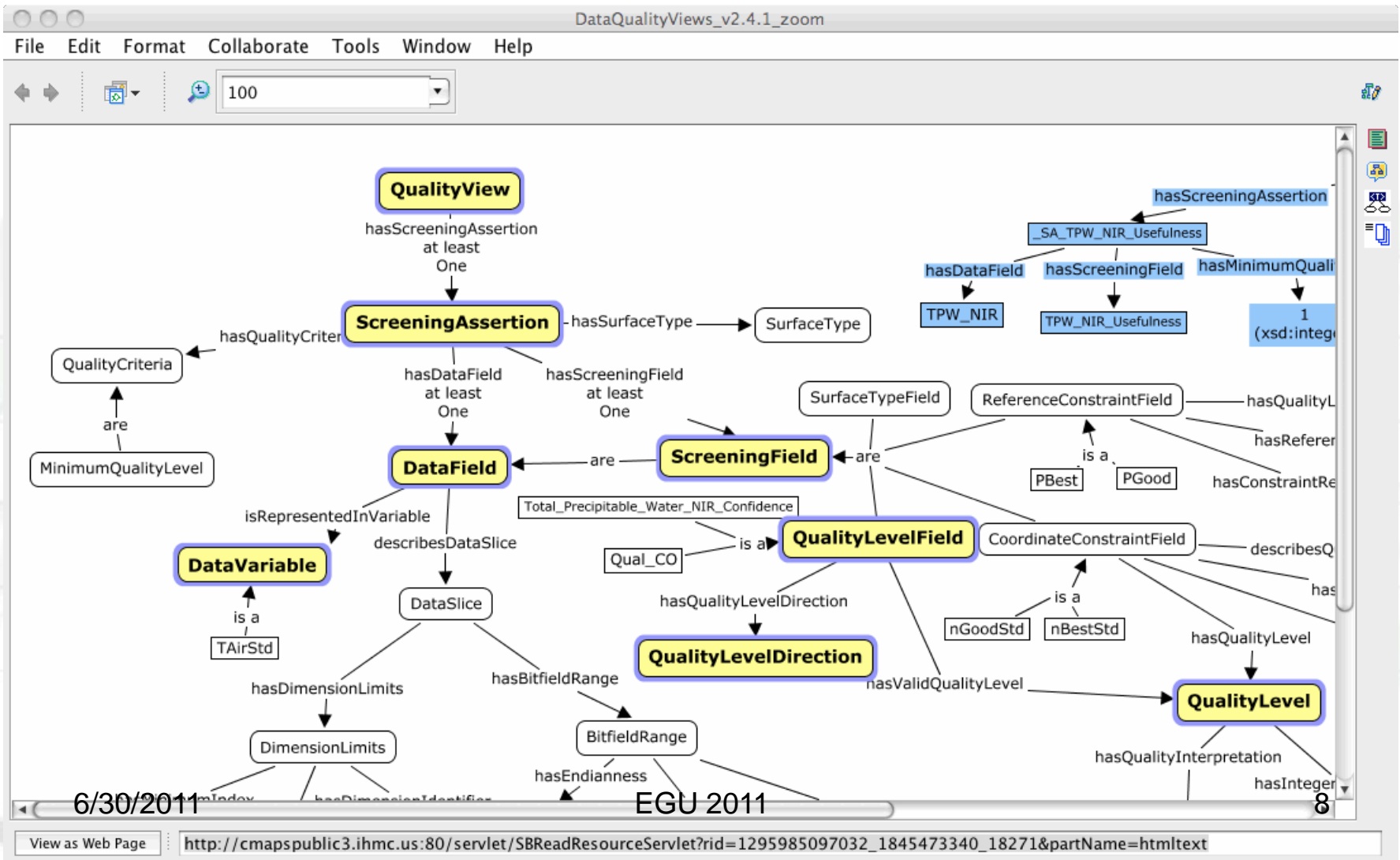


DQSS Ontology (The Whole Enchilada)





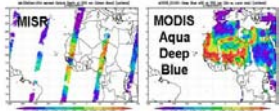
DQSS Ontology (Zoom)



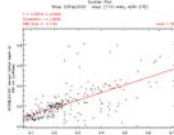


AeroStat: Online Platform for the Statistical Intercomparison of Aerosols

Explore & Visualize Level 3



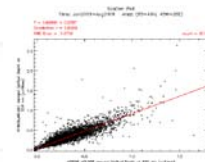
Compare Level 3



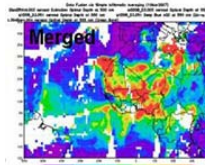
Explore & Visualize Level 2

Correct Level 2

Compare Level 2
Before and After



Merge Level 2
to new Level 3



Level 3 are too
aggregated

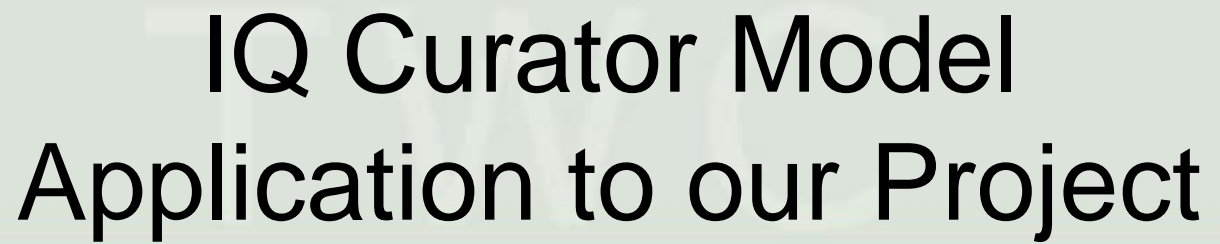
Switch to
high-res
Level 2



IQ Qurator Model

Qurator info model used our assessment of existing quality models.

- Describes a model whereby:
 - data annotated with quality annotation metadata
- QA metadata can be associated with data of varying degrees of granularity
 - ex: products, collections, arrays, specific values, etc.
 - this supports our interest in associated data with a product
- Quality evidence, a measurable quantity, provides a 'clue' into the quality
 - ex: hit-ratio, standard deviation, etc.
 - common examples associated with statistical analysis
 - often computed in QC
 - would global coverage, scatter plots, etc. fit?
- Quality assertions are domain-specific functions based on quality evidence
 - good, bad, ugly
 - No confidence, marginal, good, best
- Quality property (aka quality dimensions)
 - accuracy, completeness, currency
 - many dimensions of quality to consider, each with different evidence

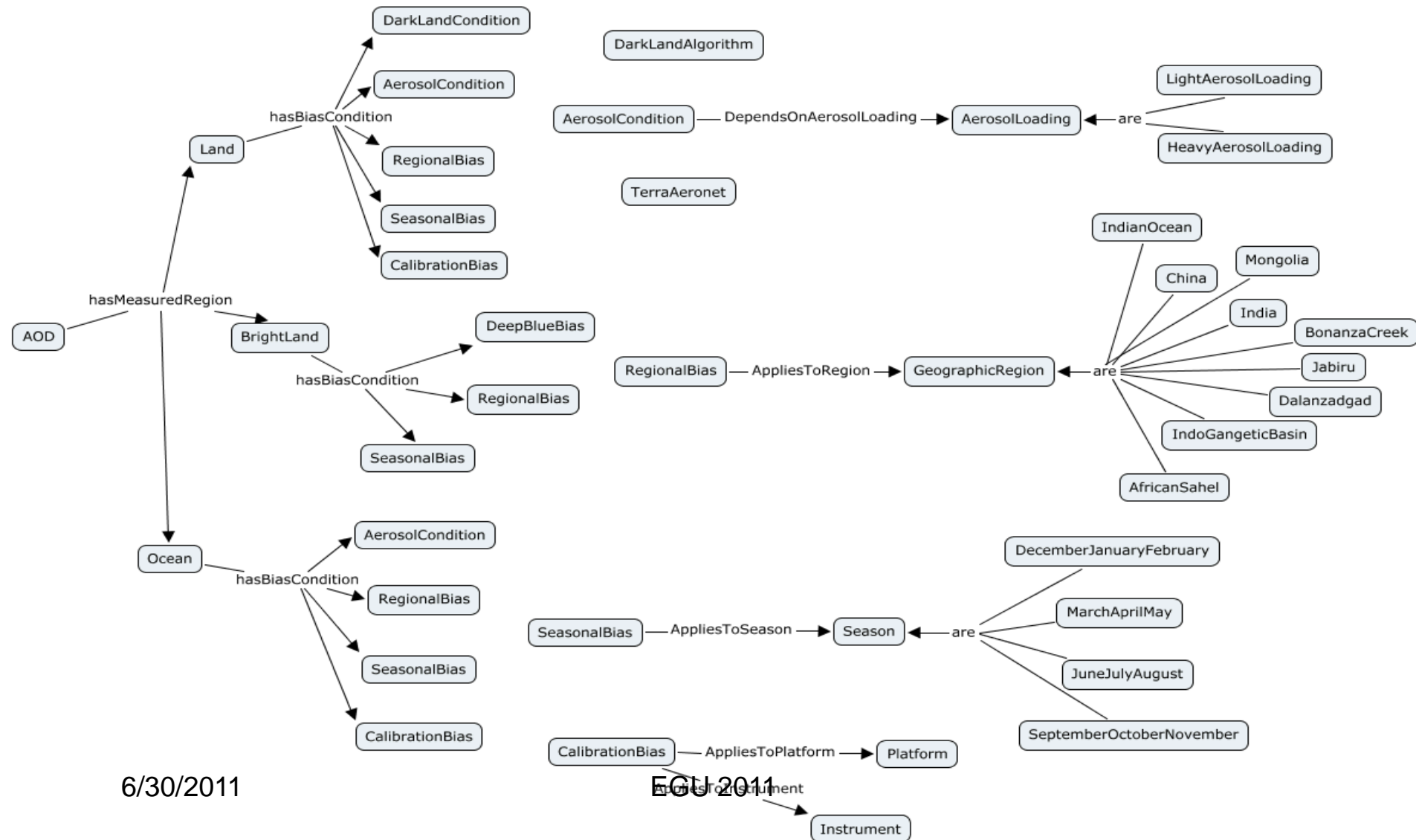


Process is general, could be used to describe report activities, QA or bias studies, research, etc.





Application to Focus Area



6/30/2011

ECU 2011



Multi-Sensor Data Synergy Advisor (MDSA)

- *Goal:* Provide science users with clear, cogent information on salient differences between data candidates for fusion, merging and intercomparison
 - Enable scientifically and statistically valid conclusions
- Develop MDSA on current missions:
 - Terra, Aqua, (maybe Aura)
- Define implications for future missions

Title: MODIS Terra C5 AOD vs. Aeronet during Aug-Oct Biomass burning-in Central Brazil,

(General) Statement: *Collection 5 MODIS AOD at 550 nm during Aug-Oct over Central South America highly over-estimates for large AOD and in non-burning season underestimates for small AOD, as compared to Aeronet; good comparisons are found at moderate AOD.*

Region & season characteristics: Central region of Brazil is mix of forest, cerrado, and pasture and known to have low AOD most of the year except during biomass burning season

(Dominating factors leading to Aerosol Estimate bias):

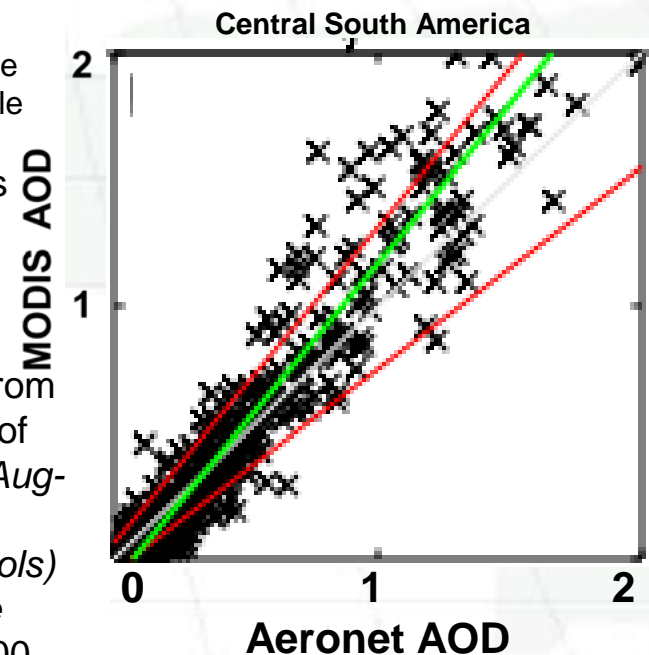
1. Large positive bias in AOD estimate during biomass burning season may be due to wrong assignment of Aerosol absorbing characteristics.
(Specific explanation) a constant Single Scattering Albedo ~ 0.91 is assigned for all seasons, while the true value is closer to ~ 0.92 - 0.93 .
[Notes or exceptions: Biomass burning regions in Southern Africa do not show as large positive bias as in this case, it may be due to different optical characteristics or single scattering albedo of smoke particles, Aeronet observations of SSA confirm this]
2. Low AOD is common in non burning season. In Low AOD cases, biases are highly dependent on lower boundary conditions. In general a negative bias is found due to uncertainty in Surface Reflectance Characterization which dominates if signal from atmospheric aerosol is low.

(Example) : Scatter plot of MODIS AOD and AOD at 550 nm vs. Aeronet from ref. (Hyer et al, 2011) (Description Caption) shows severe over-estimation of MODIS Col 5 AOD (dark target algorithm) at large AOD at 550 nm during Aug-Oct 2005-2008 over Brazil. (Constraints) Only best quality of MODIS data (Quality =3) used. Data with scattering angle > 170 deg excluded. (Symbols)

Red Lines define regions of Expected Error (EE), **Green** is the fitted slope

Results: Tolerance= 62% within EE; RMSE=0.212 ; r^2 =0.81; **Slope=1.00**

For Low AOD (<0.2) Slope=0.3. For high AOD (> 1.4) Slope=1.54





FACETS OF DATA QUALITY



Quality Control vs. Quality Assessment

- Quality Control (QC) flags in the data (assigned by the algorithm) reflect “happiness” of the retrieval algorithm, e.g., all the necessary channels indeed had data, not too many clouds, the algorithm has converged to a solution, etc.
- Quality assessment is done by analyzing the data “after the fact” through validation, intercomparison with other measurements, self-consistency, etc. It is presented as bias and uncertainty. It is rather inconsistent and can be found in papers, validation reports all over the place.



Different kinds of reported and perceived data quality

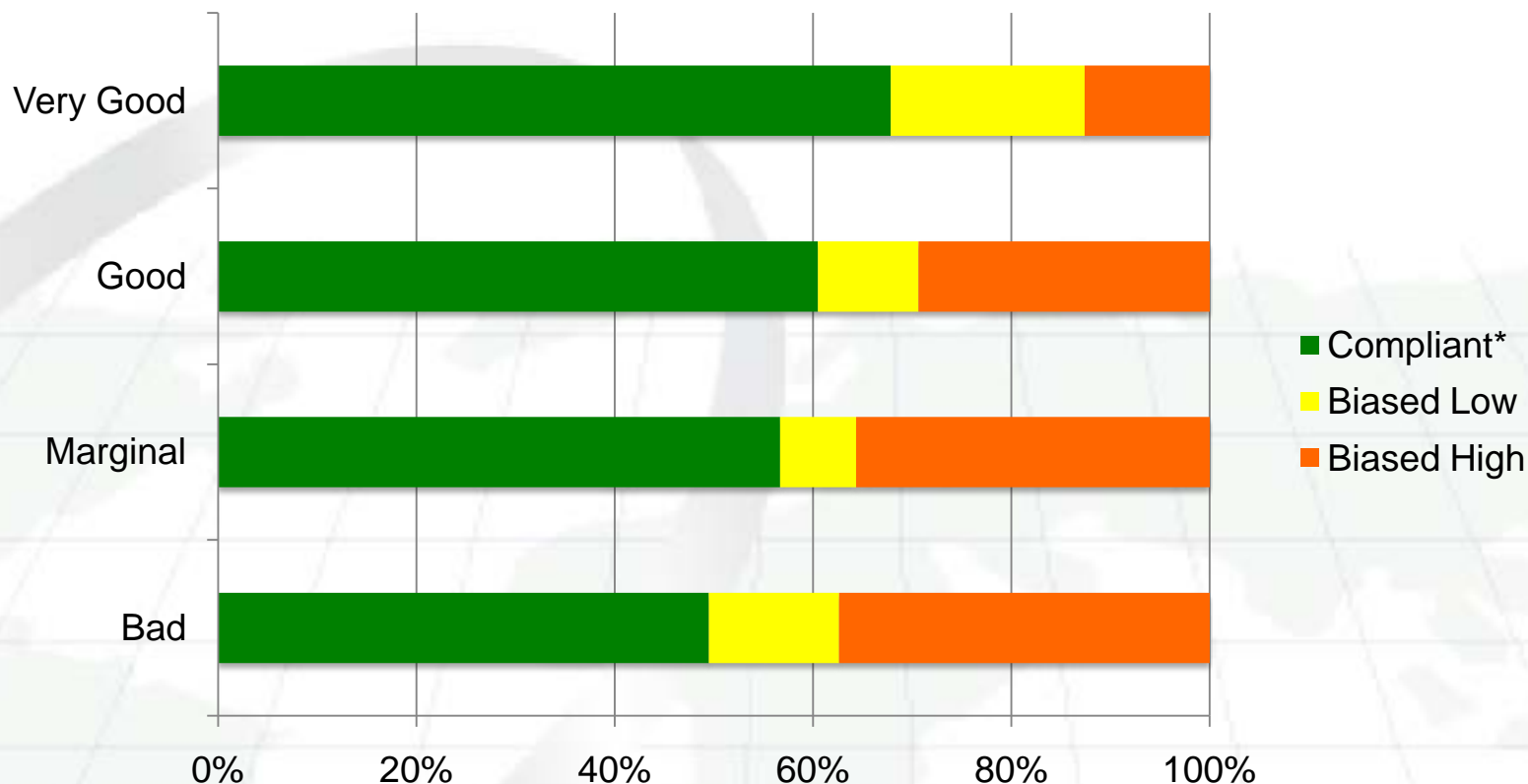
- **Pixel-level Quality** (reported): algorithmic guess at usability of data point (some say it reflects the algorithm “happiness”)
 - Granule-level Quality: statistical roll-up of Pixel-level Quality
- **Product-level Quality** (wanted/perceived): how closely the data represent the actual geophysical state
- **Record-level Quality**: how consistent and reliable the data record is across generations of measurements

Different quality types are often erroneously assumed having the same meaning

Different focus and action at these different levels to ensure Data Quality



Percent of Biased Data in MODIS Aerosols Over Land Increase as Confidence Flag Decreases



*Compliant data are within $\pm 0.05 \pm 0.2\tau_{\text{Aeronet}}$

Statistics from Hyer, E. J., Reid, J. S., and Zhang, J., 2011: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, Atmos. Meas. Tech., 4, 379-408, doi:10.5194/amt-4-379-2011



General Level 2 Pixel-Level Issues

- How to extrapolate validation knowledge about selected Level 2 pixels to the Level 2 (swath) product?
- How to harmonize terms and methods for pixel-level quality?

AIRS Quality Indicators

0 Best
1 Good
2 Do Not Use

Data Assimilation
Climatic Studies



Match up the
recommendations?

Purpose

MODIS Aerosols Confidence Flags

Ocean

Land

3 Very Good
2 Good
1 Marginal
0 Bad

3 Very Good
2 Good
1 Marginal
0 Bad

*Use these flags in order to stay
within expected error bounds*

Ocean
 $\pm 0.03 \pm 0.10 \text{ t}$

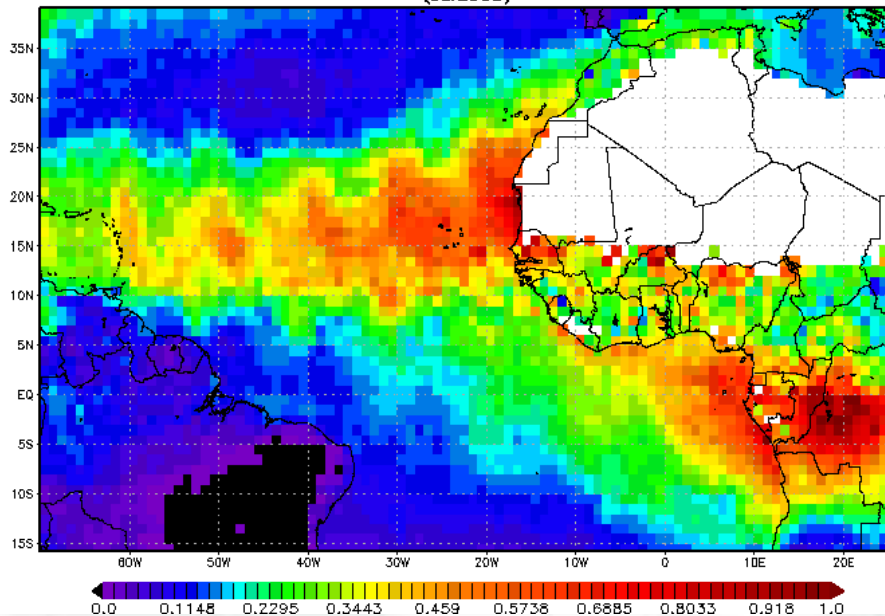
Land
 $\pm 0.05 \pm 0.15 \text{ t}$



Spatial and temporal sampling – how to quantify to make it useful for modelers?

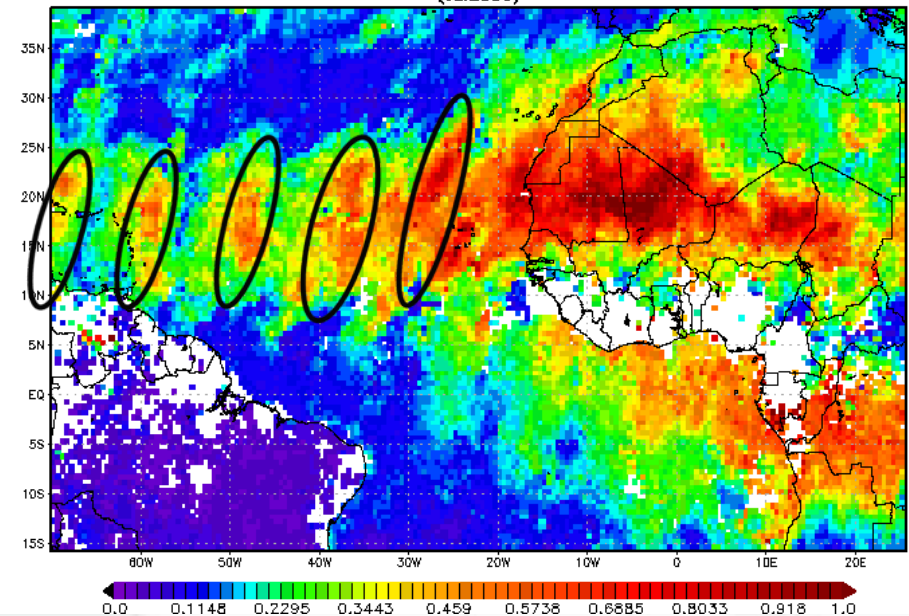
MODIS Aqua AOD July 2009

MYD08_M3.051 Aerosol Optical Depth at 550 nm [unitless]
(Jul2009)



MISR Terra AOD July 2009

MIL3MAE.004 Aerosol Optical Depth at 555 nm (Green Band) [unitless]
(Jul2009)



- Spatial sampling patterns are different for MODIS Aqua and MISR Terra:
 - “pulsating” areas over ocean are oriented differently due to different direction of orbit during day-time measurement → *Cognitive bias*



Addressing Level 3 data “quality”

- Terminology: Quality, Uncertainty, Bias, Error budget, etc.
- Quality aspects (examples):
 - Completeness:
 - *Spatial* (MODIS covers more than MISR)
 - *Temporal* (Terra mission has been longer in space than Aqua)
 - *Observing Condition* (MODIS cannot measure over sun glint while MISR can)
 - Consistency:
 - *Spatial* (e.g., not changing over sea-land boundary)
 - *Temporal* (e.g., trends, discontinuities and anomalies)
 - *Observing Condition* (e.g., exhibit variations in retrieved measurements due to the viewing conditions, such as viewing geometry or cloud fraction)
 - Representativeness:
 - Neither pixel count nor standard deviation fully express representativeness of the grid cell value



Some differences in L3 are due to difference processing

- Spatial and temporal binning (L2→L3 daily) leads to *Aggregation bias*:
 - Measurements (L2 pixels) from one or more orbits can go into a single grid cell → different within-grid variability
 - Different weighting: pixel counts, quality
 - Thresholds used, i.e., > 5 pixels
- Data aggregation (L3D → L3monthly → regional → global):
 - Weighting by pixel counts or quality
 - Thresholds used, i.e., > 2 days

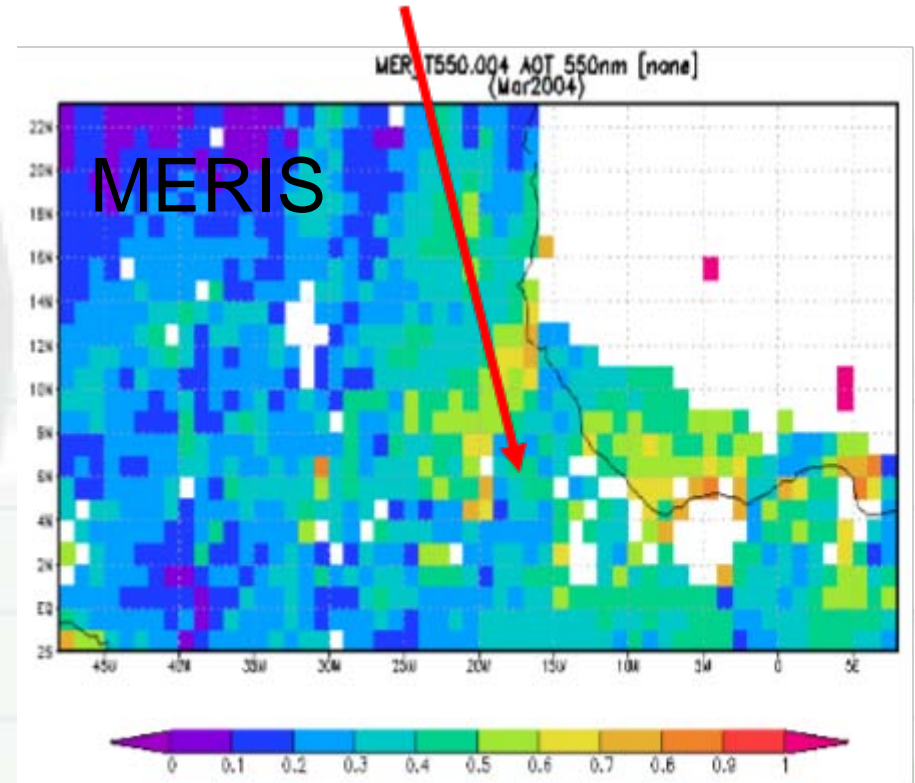
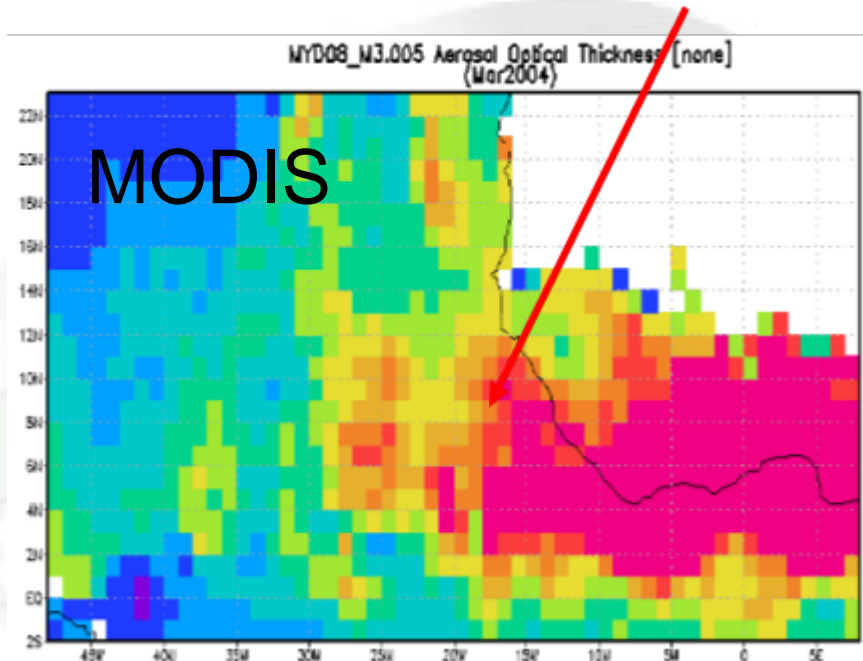
While these algorithms have been documented in ATBD, reports and papers, the typical data user is not immediately aware of how a given portion of the data has been processed, and what is the resulting impact



Case 1: MODIS vs. MERIS

Same parameter

Same space & time



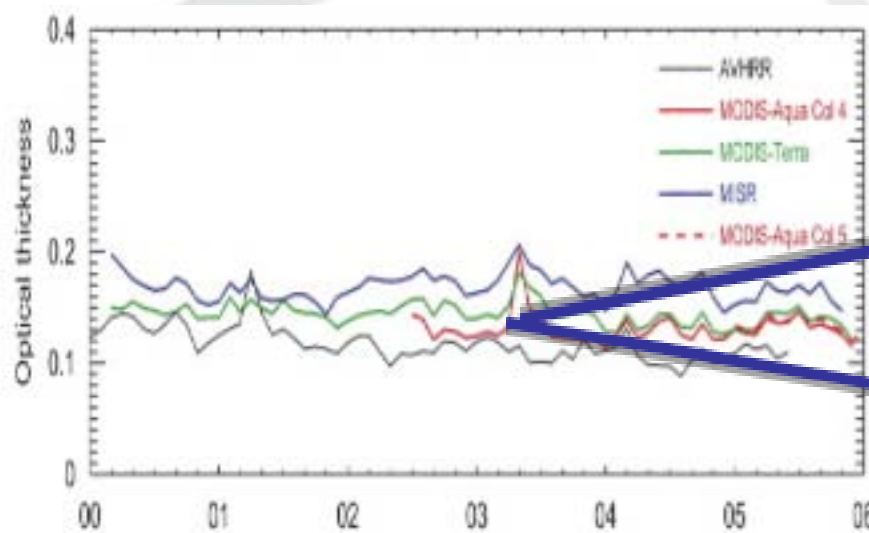
Different results - why?

A threshold used in MERIS processing effectively excludes high aerosol values. *Note: MERIS was designed primarily as an ocean-color instrument, so aerosols are "obstacles" not signal.*



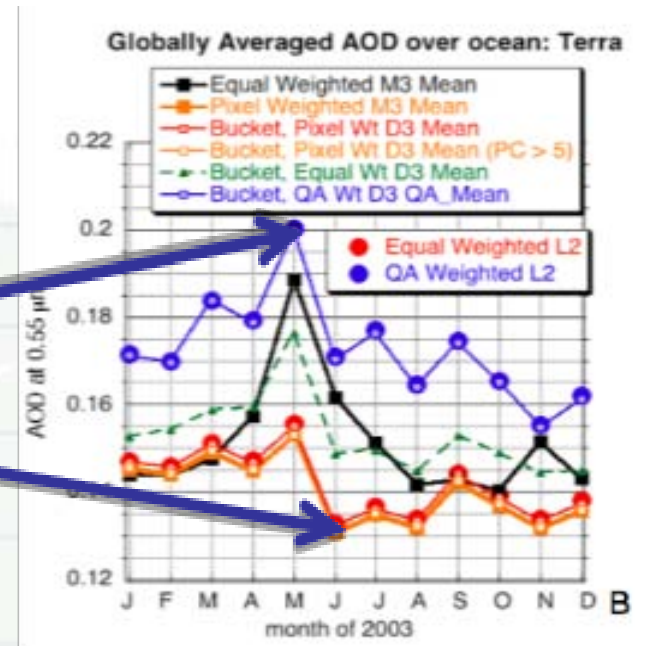
Case 2: Aggregation

AOD difference between sensors



Mishchenko et al., 2007

MODIS Terra only AOD: difference between diff. aggregations



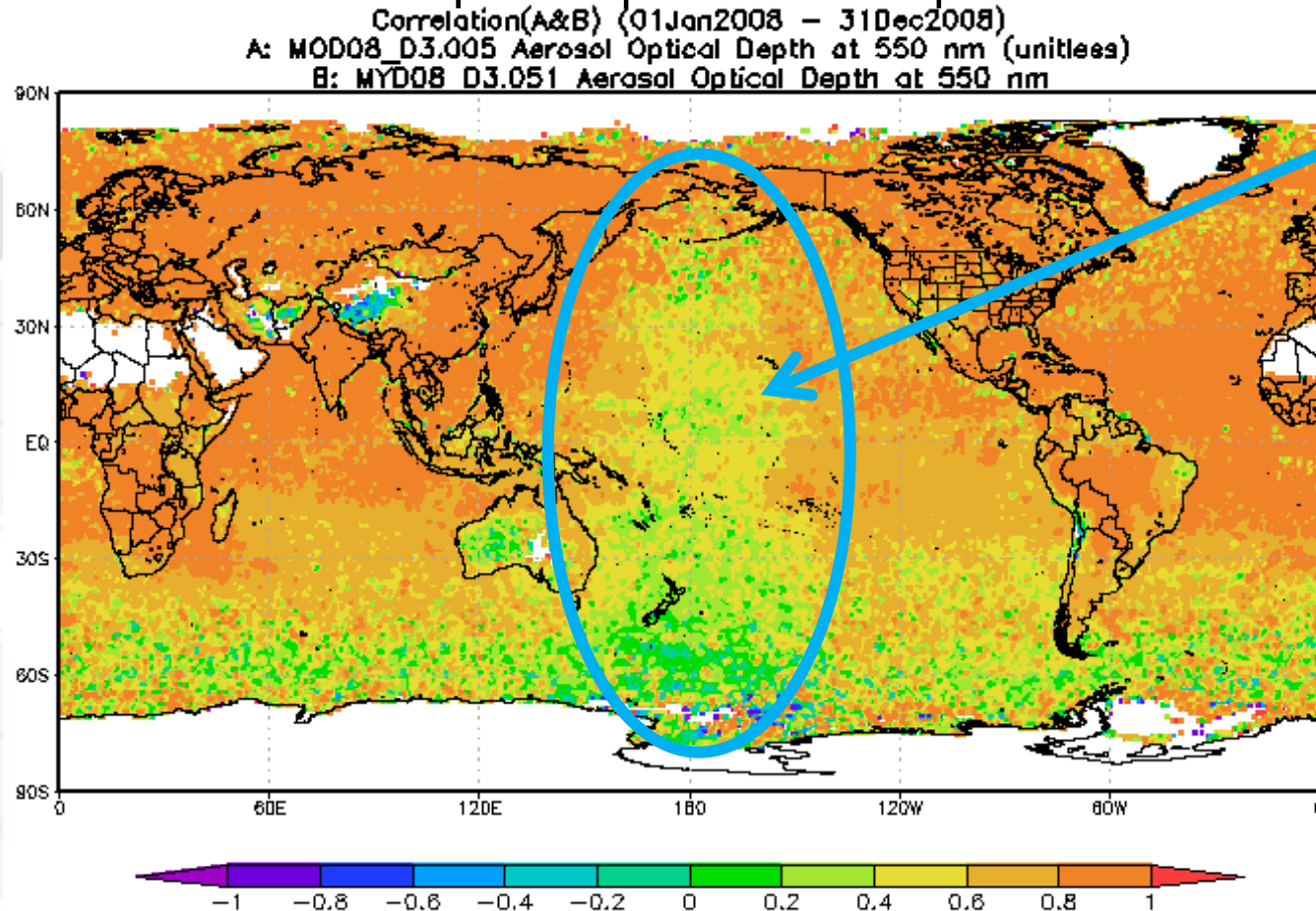
Levy, Leptoukh, et al., 2009

The AOD difference can be up to 40% due to differences in aggregation



Case 3: DataDay definition

MODIS-Terra vs. MODIS-Aqua: Map of AOD temporal correlation, 2008



MODIS Level 3 dataday definition leads to artifact in correlation



Conclusion

- Quality is very hard to characterize, different groups will focus on different and inconsistent measures of quality.
- Products with known Quality (whether good or bad quality) are more valuable than products with unknown Quality.
 - Known quality helps you correctly assess fitness-for-use
- Harmonization of data quality is even more difficult than characterizing quality of a single data product



References

- Levy, R. C., Leptoukh, G. G., Zubko, V., Gopalan, A., Kahn, R., & Remer, L. A. (2009). A critical look at deriving monthly aerosol optical depth from satellite data. *IEEE Trans. Geosci. Remote Sens.*, 47, 2942-2956.
- Zednik, S., Fox, P., & McGuinness, D. (2010). System Transparency, or How I Learned to Worry about Meaning and Love Provenance! 3rd International Provenance and Annotation Workshop, Troy, NY.
- P. Missier, S. Embury, M. Greenwood, A. Preece, and B. Jin. Quality views: capturing and exploiting the user perspective on data quality. *Procs VLDB*, 2006.
(PDF) http://users.cs.cf.ac.uk/A.D.Preece/qurator/resources/qurator_vldb2006.pdf